Evidence-Based Systematic Review of Newborn Hearing Screening Using Behavioral Audiometric Threshold as a Gold Standard*

Beth A. Prieve
Syracuse University, Syracuse, NY

Kathryn Laudin Beauchaine
Boys Town National Research Hospital, Omaha, NE

Diane Sabo
Children's Hospital of Pittsburgh and the University of Pittsburgh, Pittsburgh, PA

Tracy Schooling
American Speech-Language-Hearing Association, Rockville, MD

Brandt Culpepper
Northside Hospital, Atlanta, GA

Anne Marie Tharpe
Vanderbilt University, Nashville, TN

*This document is also included as an appendix in:

American Speech-Language-Hearing Association. (2013). *Expert panel recommendations on newborn hearing screening*. Available from: www.asha.org.

Abstract

Purpose: Although there are several evidence-based systematic reviews (EBSRs) that provide evidence that universal newborn hearing screening (UNHS) identifies hearing loss at an early age and results in better language outcomes, there is a lack of EBSRs that evaluate the effectiveness of hearing screening tools that can guide decisions for developing best-practice protocols. The goal of this EBSR was to evaluate the literature regarding the comparative effectiveness of physiologic screening tools (i.e., otoacoustic emissions [OAEs], auditory brainstem response [ABR], or auditory steady-state response [ASSR]) for identifying permanent hearing loss and the "gold standard," behavioral audiometric threshold testing. Criteria for study inclusion were that at least a portion of infants who passed the newborn hearing screening were included in follow-up and that behavioral audiometric thresholds were measured.

Method: The literature was systematically searched using 18 electronic databases. A total of 12 studies that addressed five questions were identified for inclusion. Initially, two reviewers evaluated the studies; a review panel further analyzed and discussed all included studies. Positive and negative likelihood ratios (LR+ and LR-) were calculated based on sensitivity and specificity for groups of infants who returned for behavioral testing.

Results: The LR+ from the studies ranged from 1.75 to 87.9, with 47% of the LR+ greater than 5—interpreted as having at least a moderate likelihood of hearing loss. Of interest was that the majority of the studies were published before 2000. Also noteworthy was the fact that none of the studies included behavioral threshold testing of babies in the well-infant nursery (WIN) who passed newborn hearing screening and did not have risk factors for hearing loss.

Conclusion: All of the screening studies indicated an increase in the likelihood of hearing loss based on a failed newborn hearing screening, with approximately half indicating a moderate or greater increase in the likelihood of hearing loss. No studies that met inclusion criteria employed currently used screening techniques and equipment, thereby limiting their usefulness to guide recommendations for best-practice protocols. Research is needed on the effectiveness of different protocols and methods for identifying hearing loss in populations of infants in well-baby and intensive care units.

Key words: systematic review, newborns, hearing screening, OAEs, ABR

Early hearing detection and intervention (EHDI) programs exist in all 50 states and territories of the United States (National Center for Hearing Assessment and Management [NCHAM], 2012). Of all of the conditions for which we screen at birth, congenital hearing loss has the highest prevalence, with hearing loss having a higher prevalence than all other screened conditions combined (Jacobson & Jacobson, 2004). Early diagnosis and management of hearing loss in infants promotes age-appropriate speech and language outcomes (Moeller, 2000; Sininger, Grimes, & Christensen, 2010; Yoshinaga-Itano, Sedey, Coulter, & Mehl, 1998). Enhancing newborn hearing screening and follow-up procedures is critical for optimizing the outcomes for children with hearing loss.

Current clinical practice in the United States and in many other countries around the world is to evaluate health care programs using evidence-based practice (EBP). One part of EPB is the conduct of an EBSR in which clinical questions are asked and supporting evidence compiled to address those questions. Several EBSRs have been conducted that support the screening of every newborn for hearing loss. The questions in each EBSR differ, as do the outcomes measures chosen to evaluate screening effectiveness.

An early EBSR performed by a team in Great Britain (Davis et al., 1997) commenced because it was becoming apparent that programs used at the time (Health Visitor Distraction Test [HVDT]) to screen children at 7–8 months of age in the home did not adequately screen for hearing loss. Five questions were asked.

1. What is the current epidemiology of permanent childhood hearing loss (PCHL) in the United Kingdom (UK)?
2. What are the outcome benefits of early identification of PCHL?
3. What is the current practice in the UK for screening hearing loss at birth and at school entry?
4. What are the likely costs associated with current screening programs?
5. What is the effectiveness of universal neonatal, targeted neonate, and HVDT screening approaches?

The EBSR revealed that

1. approximately 840 children a year are born in the UK with significant PCHL—400 would be missed by $1^1/_2$ years of age and 200 would be missed by $3^1/_2$ years of age;
2. children with PCHI identified later are at risk for delay of communication skills;
3. practices at the time varied among regions;
4. there was poor sensitivity and relatively poor specificity for the HVDT with relatively low yield;
5. median age of identification was 12–20 months.

Neonate screening showed high sensitivity and reasonably high specificity. UNHS at the time was not done routinely. Cost per child was lower for universal screening than for HVDT. They concluded that UNHS had a lower running cost per child detected than HDVT. Coverage was greater than 90%, and specificity was about 95%. Sensitivity was high on a small neonatal sample.

The first EBSR of the U.S. Preventive Service Task Force (USPSTF, 2001) asked the following questions.

1. Can UNHs accurately diagnose moderate-to-profound sensorineural hearing impairment?
2. In UNHS programs, how many children are identified and treated early?
3. Does identification and treatment prior to age 6 months improve language and communication?
4. What are the potential adverse effects of screening and of early treatment?

The compiled evidence supported that UNHS resulted in earlier identification of hearing loss than was suggested by the previous literature for which there was no UNHS, but could not find high-quality evidence that UNHS programs resulted in better language outcomes than resulted from identification of loss later in life. The evidence also could not adequately address the adverse effects of screening and early treatment. USPSTF conducted a second EBSR (Nelson, Bougatsos, & Nygren, 2008) that examined evidence for "(1) the efficacy of UNHS in improving the initiation of treatment by 6 months of age for

average- and high-risk infants compared with targeted screening (2) the efficacy of treatment on language and communication outcomes if started before 6 months of age for those infants not identified by targeted screening and (3) the harms of universal newborn hearing screening" (USPSTF, 2008). USPSTF found that infants screened in a UNHS program had earlier detection, intervention, and initiation of treatment than did a group who were not screened. The task force also concluded that there was good evidence that children who had bilateral PCHL and had diagnostic confirmation by 9 months of age had higher receptive language outcomes at 8 years of age than those who were not screened. However, they found no studies that directly compared initiation of treatment—via targeted screening versus UNHS—for infants at average and high risk for hearing loss. Finally, they found conflicting results regarding the degree of stress experienced by parents whose babies did not pass the screening as compared with those whose infants passed.

An EBSR commissioned by the German Federal Joint Committee investigated the benefits and harms of identifying hearing loss in newborns through mass screening programs (German Institute for Quality and Efficiency in Health Care [IQWiG], 2007). Their EBSR team addressed three research areas: (1) the effectiveness of the screening programs in terms of different times of screening, screening for different severities of hearing loss, and other differences; (2) the effectiveness of treatment at different ages in the child's life; and (3) the sensitivity/specificity of OAE screening followed by ABR screening. Similar to conclusions of USPSTF (2001, 2008), they found evidence to support that UNHS resulted in earlier identification of congenital hearing loss with better outcomes than outcomes for newborns who did not participate in UNHS programs. Furthermore, they found an indication that those who were identified with hearing loss earlier had more favorable language development at 3 and 8 years of age relative to those identified later. To investigate the accuracy of using OAE and ABR screening in combination to identify hearing loss, they included infants who had been screened under 1 year of age and whose hearing was evaluated at later ages using "any sort of reference test" (Table 3, p. 15). Their EBSR results relied heavily on the work by the Wessex Universal Neonatal Hearing Screening Trial Group (Kennedy, McCann, Campbell, Kimm, & Thornton, 2005; Kennedy, 1999). In the Wessex study, the reference test

was HVDT, which is conducted by a visiting nurse in the baby's home. They found that test accuracy was favorable when using a two-technology screening approach, where OAE screening was followed by ABR screening (91.7% sensitivity and 98.5% specificity). While there was some evidence that the combination of tests showed good test performance, they did not address each test measure's performance without the other. The HVDT reference standard is an excellent method to obtain behavioral information on a large number of infants who pass and fail UNHS; however, it is not considered a diagnostic hearing test because it is conducted in the home rather than in a clinical setting and completed by health visitors, rather than audiologists.

These EBSRs provided strong support for UNHS in that the screening lowered the age of identification of PCHL and resulted in improved language outcomes compared with outcomes for those not being screened. However, the questions were not aimed to guide which test measures and criteria should be recommended for UNHS. The American-Speech-Language-Hearing Association (ASHA) organized a working group to develop guidelines for hearing screening, the first step of which is to conduct an EBSR. The working group identified important issues regarding protocols and personnel that are critical for guiding staff to operate the most effective NBHS programs. One of the most basic questions posed was the effectiveness of ASSR, OAE, and ABR; the latter two screening tools are recommended by the Joint Committee on Infant Hearing (JCIH, 2007) for identifying PCHL. *Hearing loss* was defined in the broadest sense to encompass the entire auditory pathway and represent what is perceived by an individual. Therefore, the reference standard of behavioral audiometric hearing tests was chosen. For infants, however, this reference standard is separated in time when used to evaluate hearing, because an accurate measure of behavioral audiometric threshold cannot be performed until the infant is at least 6 months developmental age.

The purpose of the current manuscript is to describe the results for five key questions posed:

1. For infants birth through 6 months, what is the effectiveness (sensitivity/specificity) of OAEs in identifying children with hearing loss?

2. For infants birth through 6 months, what is the effectiveness (sensitivity/specificity) of ABR in identifying children with hearing loss?
3. For infants birth through 6 months, what is the effectiveness (sensitivity/specificity) of ASSR on identifying children with hearing loss?
4. For infants birth through 6 months, what is the effectiveness of various OAE stimulus parameters (level and frequency) in identifying children with hearing loss?
5. For infants birth through 6 months, what is the effectiveness of various ABR stimulus levels in identifying children with hearing loss?

A requirement of the studies was that at least a portion of infants passing newborn hearing screening needed to return for behavioral audiologic threshold testing.

Methods

A systematic search of the literature was conducted and studies were considered for inclusion if they were published in a peer-reviewed journal (as classified by *Ulrich's Periodicals Directory*) from 1975 to 2008, were written in English, and contained original data addressing one or more questions. Studies were excluded if an appropriate reference standard was not used to determine the accuracy of the initial screening results. The reference standard was behavioral audiometric thresholds. The working group included studies that employed a conditioned audiometry technique that uses a response reinforcer, such as Visual Reinforcement Audiometry, Conditioned Play Audiometry, Conditioned Orienting Response, or Visual Reinforcement Orientating Audiometry. Studies were also included that described behavioral audiometric testing (without reference to response reinforcement), if details were provided about hearing threshold levels that were used to classify an ear or participant as having hearing loss or if descriptions were included of eventually obtaining accurate behavioral audiometric thresholds on the infants. Studies were excluded if only electrophysiological examinations (e.g., ABR, OAEs, ASSR) or less-controlled behavioral examinations of hearing (i.e., behavioral observation audiometry, HVDT) were used. No criterion was set for transducer (e.g., speaker or earphones) or stimulus (e.g., tone or speech) type. A second criterion was

that studies were included only if an appropriate reference standard was administered to at least a subset of those who passed, as well as those who failed, the initial screening. Finally, studies were excluded if they did not report or supply sufficient data to calculate both the sensitivity and specificity of the screening tool.

Eighteen electronic databases were searched using a series of key words and expanded search terms related to infants, hearing screening technology, and diagnostic accuracy (Appendix I). Electronic databases included the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Library, Communication and Mass Media Complete, Education Abstracts, Education Resources Information Center (ERIC), Evidence-Based Medicine Guidelines, Health Source: Nursing, HighWire Press, Linguistics Language Behaviour Abstracts, PsycINFO, Psychology and Behavioral Sciences Collection, PsycArticles, PubMed, Science Citation Index, ScienceDirect, Social Science Citation Index, SUMSearch, and Turning Research into Practice (TRIP) Database. A supplemental search of several key websites identified resources from the Medical Research Council's Institute of Hearing Research, National Center on Birth Defects and Developmental Disabilities (EHDI program-related publications), and the Otoacoustic Emissions Portal Zone. Additionally, a manual search of references from all relevant articles was completed.

As displayed in Figure 1, a total of 1,024 citations were identified. Initially, two reviewers, blinded from one another's results, reviewed each abstract and identified 236 citations as preliminarily meeting the inclusion criteria with 87% agreement. Of those, 223 were subsequently excluded by these two reviewers, sometimes in consultation with the larger review panel (consisting of the five co-authors), for a total of 13 studies for inclusion, with 12 addressing the questions described in this manuscript.

The two initial reviewers, still blinded to one another's results, assessed studies for methodological quality. Studies were assessed in the following areas: study design, appropriateness of reference standard, selection/recruitment, assessor blinding, participant description, avoidance of verification bias, LRs, and follow-up. Each study received a point for each marker meeting the highest level of quality (Table 1); a final score was derived from the total number of indicators that met the highest level of quality. Given that the inclusion criteria of this EBSR incorporated two of the quality appraisal indicators (i.e., appropriateness

of reference standard and LRs calculated from sensitivity and specificity values), each included study had a minimum quality score of 2.

Each critical appraisal was reviewed by at least one member of the evidence panel, who also completed the data extraction of key study variables (i.e., participant description, screening tools, stimulus parameters, reference standard description). Agreement between the initial and panel reviewers was greater than 98%, and any discrepancies in ratings were resolved via consensus by the full panel.

A requirement of EBSRs conducted through ASHA is that LRs need to be calculated. In order to do so, measures of sensitivity and specificity are needed. In order to calculate sensitivity and specificity, only the babies who had been followed up with behavioral hearing testing were included in the calculation from each study. This method differs from most other NBHS reports in which sensitivity is calculated based on the total number of ears or infants who passed the screening, assuming that all passed infants had normal hearing. Because our calculations were based on the number of infants who were behaviorally tested, the calculated sensitivity and specificity for most cases differ from those reported by the authors of the study. Sensitivity and specificity were calculated using the following equations:

Sensitivity = # of infants or ears that failed the hearing screening and had permanent childhood hearing loss (PCHL) by the diagnostic test/total # of infants or ears who had PCHL as measured by behavioral audiometric thresholds

Specificity = # of infants or ears who passed the hearing screening/# of infants or ears that had normal hearing by behavioral audiometric thresholds

Positive and negative LRs were calculated from study sensitivity and specificity using the formulas:

Positive likelihood ratio (LR +) = sensitivity/1− specificity

Negative likelihood ratios (LR -) = 1− sensitivity/specificity

LRs reflect the likelihood of having a disorder based on a positive or negative screening result (Dollaghan, 2007). LR+ values indicate the likelihood that the ear being tested has hearing loss. LR- is the decreased likelihood that an ear has no hearing loss (normal hearing). Table 2 provides the interpretation of LR scores used in this EBSR.

Results

The literature search produced 12 studies addressing newborn hearing screening that measured behavioral audiometric threshold in at least a subset of the infants who passed the hearing screen (Apostolopoulos, Psarommatis, Tsakanikos, Dellagrammatikas, & Douniadakis, 1999; Ari-Even Roth et al., 2008; Desai et al., 1997; Durieux-Smith, Picton, Bernard, MacMurray, & Goodman, 1991; Gill, Gosling, Kelly, Walker, & Wooderson, 1998; Norton et al., 2000; Savio, Perez-Abalo, Gaya, Hernandez, & Mijares, 2006; Shimizu et al., 1990; Smyth, Scott, & Tudehope, 1990; Stevens et al., 1990; Swigonski, Shallop, Bull, & Lemons, 1987; Watkin, Baldwin, & McEnery, 1991).

Table 3 provides a description of participants for each study: the number of infants and/or ears screened, the age at time of screening, the type of nursery in which the infants received care, the number of infants and/or ears received follow-up, age at follow-up, and a description of the hearing loss. Most studies included only infants cared for in the neonatal intensive care unit (NICU), with the possible exceptions of the Norton et al. (2000) and Ari-Even Roth et al. (2008) studies, which did not state whether infants were cared for in the NICU or the well-infant nursery (WIN). In the study by Norton and colleagues (2000), 7,179 infants were screened before hospital discharge;

2,348 of those infants were in the WIN, including 353 who had risk indicators for hearing loss (JCIH, 1994). However, only infants cared for in the NICU and infants with risk indicators cared for in the WIN (n = 4,911) were targeted for follow-up behavioral testing. Also noted, the number of infants followed up for behavioral testing was considerably lower than the number screened, and calculations for the current investigation were based on the number who had a behavioral hearing test, not the number screened.

Table 4 summarizes the quality of the appraisal for each study based on the quality indicators listed in Table 1. The gray areas highlight the study factors that met the highest level of quality. As noted earlier, because of the EBSR inclusion criteria, two quality indicators—reference standard and the ability to compute LRs—represented the highest level for all studies. Most studies also had the highest-level quality indicator in study design and avoidance of verification bias, with one study in each quality

category not meeting the highest standard. No study had the highest quality indicator for the subjects being similar to the population being studied, because no study included infants cared for in the WIN in the follow-up who did not also have a risk factor for hearing loss (Norton et al., 2000). Most studies were not rated highly in the percentage of subjects who completed follow-up, with most reporting that more than 20% of subjects were lost to follow-up.

Table 5 presents key variables (e.g., stimulus parameters, reference standards, hearing loss definitions) as well as the sensitivity/specificity and LRs for each of the studies addressing the question: For infants birth through 6 months, what is the effectiveness (sensitivity/specificity) of OAEs in identifying children with hearing loss? Two different types of OAEs (transient and distortion product), both recommended by JCIH (2007), were used. Stimulus parameters and response criteria for a "pass" varied among studies. Although all studies reported using behavioral measures as the follow-up test, the amount of detail provided by the studies varied widely. Some only reported that a behavioral technique was used (Gill et al., 1998), while others provided extensive detail on the behavioral procedure (e.g., behavioral methods for Norton et al., 2000, are described in Widen et al., 2000). Likewise, some reported details on type of transducer used to deliver the stimuli (Norton et al., 2000), whereas others did not (Watkin et al., 1991). The sensitivities ranged from 55% (Stevens et al., 1990) to 100% (Ari-Even Roth et al., 2008). The specificities ranged from 71% (Ari-Even Roth et al., 2008) to 91% (Apostolopoulos et al., 1999). Positive LR values ranged from a small (LR+ = 3.1) to a large (LR+ = 10.21) increase in the likelihood that hearing loss was present. This variability was also noted in LR- values, which ranged from a minimal (LR- = 0.54) to a large (LR- = 0) decrease in the probability of a disorder. .

Table 5 also includes data addressing the question: For infants birth through 6 months, what is the effectiveness of various OAE stimulus parameters (level and frequency) in identifying children with hearing loss? Norton et al. (2000) was the only study that addressed the question using DPOAEs. Two stimulus levels, f1/f2 levels of 65/50 dB SPL and 75/75 dB SPL, were run on each ear; the respective sensitivity values were 88% and 78%, and specificity values were 83% and 82%. The LR+ for the level of

65/50 (LR+ = 5.2) was higher than that for stimulus levels of 75/75 (LR+ = 4.3). The LR-s for the two levels were 0.14 and 0.27 for the 65/50 and 75/75 conditions, respectively.

Table 6 provides information addressing the question: For infants birth through 6 months, what is the effectiveness (sensitivity/specificity) of ABR in identifying children with hearing loss? For the ABR studies, sensitivity ranged from 42% (Desai et al., 1997) to 100% (Shimizu et al., 1990; Smyth et al., 1990; Swigonski et al., 1987; Watkin et al., 1991). Specificity ranged from 70% (Swigonski et al., 1987) to 100% (Durieux-Smith et al., 1991). The LRs for hearing loss ranged from minimal (LR+ = 1.75; LR- = 0.84) to large (LR+ = 87.86; LR- = 0). The last entry in Table 6 includes a description of the one study that addressed the question: For infants birth through 6 months, what is the effectiveness (sensitivity/specificity) of ASSR in identifying children with hearing loss? As indicated, the sensitivity (100%), specificity (71.3%) and LRs (LR+ = 3.48; LR- = 0) were within the range found for other screening technologies. Additional ASSR screening studies are needed to determine if these LRs are representative.

Figure 2 illustrates the LR+ and LR- for the OAE, ABR, and ASSR studies shown in Tables 5 and 7. All studies except one (Desai et al., 1997) had LR+ of 2 or higher. The LR+ and LR- for this study are noted by a diamond symbol, because the population studied included only infants treated with extracorporeal membrane oxygenation (ECMO), a procedure that is known to be associated with late-onset hearing loss (Fligor et al., 2005). Excluding the 1997 study by Desai et al., there were 19 calculable LR+ and one LR+ that was not calculable. There were nine ABR studies, with two studies providing multiple LRs based on ability to identify different types and severities of hearing loss (Durieux-Smith et al., 1991; Stevens et al., 1990). There were five OAE studies, with two studies (Norton et al., 2000; Stevens et al., 1990) providing more than one LR.

All LR+s for newborn screening indicated an increased likelihood of hearing loss. There were two LR+s indicating a slight likelihood, 10 LRs that indicated a small likelihood (2–5), and 7 LRs indicating a moderate likelihood (>5-9). Two LR+s were greater than 10, which is interpreted as a large and conclusive increase in the likelihood of hearing loss. LR- indicates decrease in the likelihood of the

disorder. In seven studies, the LR- indicated large and almost conclusive decrease in the likelihood of hearing loss and five LR-s indicated moderate decrease in the likelihood of the disorder. Five LR-s could be interpreted as a small decrease in the likelihood of the disorder, and four indicated minimal decrease of the likelihood of the disorder. LR+ and LR- were not computed for aggregate data, because definitions for hearing loss among studies varied and some studies provided more than one calculation of LR.

Table 7 reports information for the only study found that addressed the question: For infants birth through 6 months, what is the effectiveness of various OAE response criteria (frequencies and levels) in identifying children with hearing loss? Norton et al. (2000) had data addressing the LR using different SNR criteria. Table 7 lists three SNRs that were arbitrarily chosen and the resulting sensitivity, specificity, and LRs calculated for TEOAEs, DPOAEs with primary stimuli presented at 65/50 dB SPL, and DPOAEs with primary stimuli presented at 75/75 dB SPL. In addition, two definitions of hearing loss were given. It can be seen that the sensitivity increased with higher SNR criteria, but expectedly, specificity decreased, resulting in decreasing LR+s with increasing SNR. In general, LR-s across SNRs were more similar, with most being interpreted as moderate-to-conclusive decreases in likelihood of the disorder.

Discussion

The questions posed by this EBSR probed the effectiveness of screening tools for detecting PCHL using the behavioral audiometric threshold as a reference-standard. Additionally, studies were included only if at least a portion of infants passing the screening returned for audiometric threshold testing, so that LRs could be computed. The authors sought to uncover measures and criteria that would provide useful information for the development of EDHI program guidelines in the United States.

An overall finding of the EBSR is that few studies were found that met the inclusion criteria of requiring confirmation of hearing loss with behavioral audiometry in at least a portion of infants who passed and infants who failed newborn hearing screening. Most studies that met the criteria were published between 1980 and 2000, and many were from countries other than the United States. This situation limits the EBSR from being generalized to the current status of UNHS in the United States,

because there have been significant advances in technology and care of newborns since that time period. Only one article (Ari Evan-Roth et al., 2008) was published recently enough to have included currently used screening equipment and techniques. Because of the age of the studies, most that used ABR used conventional ABR rather than automated ABR, which now is commonly used in nurseries. This situation accounts for differences in transducer and calibration-response criteria and, thus, differing pass/refer results. The majority of OAE studies used equipment still in use today, but most likely with different response criteria methods.

Because many of the studies were from countries other than the United States, it is possible that NICU populations in these studies may be different from those in the United States. Other countries have different health care systems and ethic/racial distributions, and their population's genetic predisposition to hearing loss varies. Also, procedures and definitions related to the NICU, infant populations, screening protocols, and hearing loss may vary by country.

Additionally, no studies were found that included babies cared for in the WIN in their follow-up procedures, except for infants who had risk indicators for hearing loss (Norton et al., 2000). The most obvious reason why no newborns cared for in the WIN were included relate to cost: It is extremely costly to track and behaviorally test all infants who passed newborn hearing screening. The question must be raised whether identification of hearing loss in infants cared for in the NICU is representative of identification of babies born in WINs. There are differences in the prevalence of hearing loss between infants cared for in the NICU compared with infants in the WIN (Prieve 2000; Prieve & Stevens, 2000). There is a greater possibility of a NICU infant having a risk indicator for hearing loss and, in addition, the causes of hearing loss encompass a wider range of etiologies. Although the prevalence of having a risk indicator for hearing loss is different for infants cared for in the NICU as compared with those in WI nurseries, it might be safe to assume that infants born with permanent congenital hearing loss can be identified with equal accuracy in NICU and WIN populations. However, it is not likely that the development of late-onset hearing loss will be the same between groups. For example, Desai et al. (1997)

found that ABR screening did not accurately predict later hearing loss in infants treated with ECMO, a procedure not conducted on infants in WINs.

This EBSR also highlights several factors that affected results of the UNHS programs included in this EBSR that should be considered in future research. The first factor deals with hearing loss itself and the relationship between hearing screening and hearing loss diagnosis. The reference standard against which the hearing screening test was assessed in this EBSR, a behavioral hearing test, cannot be done until the infant is at least 6 months of age. During the time between the screening and the behavioral confirmation of hearing loss, there may be improvement in hearing status (i.e., resolution of transient middle ear or Eustachian dysfunction) or changes in hearing status (i.e., late-onset/progressive permanent hearing loss or worsening of middle ear dysfunction) .

In the included studies, some of the behavioral testing was completed much later than the newborn hearing screening—for example, at 18–36 months of age. In addition, studies were included in which behavioral audiometric testing was performed in a sound field rather than through an earphone, which would have left unilateral hearing loss undiagnosed. The behavioral audiometric threshold as the reference standard was chosen for this EBSR, because that screen was judged to be critical to determine how a child *hears* and responds to sounds rather than a physiological measure such as ABR, which is predictive of behavioral thresholds. It is possible that LRs would be higher if screening results were compared to diagnostic audiologic testing by frequency-specific methods, such as tone burst ABR or ASSR by 3 months of age, which is the current standard of care for diagnosis/confirmation of hearing loss. A possible limitation for this question is that, in order to calculate LRs, infants passing and failing UNHS would need to undergo these evaluations.

A second factor affecting the included studies is that infants who had OAE screening only may have had auditory neuropathy spectrum disorder (ANSD), which would have been missed. This has been addressed by JCIH (2007), which recommends that ABR be used for the NICU population. Regardless, the fact that some of the studies used OAEs for infants cared for in the NICU could have an impact on sensitivity and the LRs.

Finally, the sensitivity and specificity upon which LRs were calculated represent samples of the larger population. Although it is assumed that the subsets represent the entire population, this may not be accurate, as there were often no criteria stated for only following a subset of the infants. That being said, most of the calculated LRs did not vary widely among studies.

Despite the limitations of the studies included in the EBSR, it can be concluded that a failed OAE and ABR hearing screening at birth is associated with an increased likelihood of PCHL. Approximately half of the studies indicated at least a moderate likelihood of hearing loss. Additionally, the majority of studies had similar LR+s and LR-s, suggesting that OAE and ABR technologies provide similar results. Although this EBSR used relatively strict criteria that limited the inclusion of mostly older, published articles, the conclusions are consistent with a recent study, which found that approximately 50% of children with hearing loss entering kindergarten had passed their newborn hearing screening (Watkin & Baldwin, 2011). The children entering kindergarten had undergone UNHS using TEOAES, followed by ABR after a failed TEOAE screen. The authors found that, although some of the children who had hearing loss at kindergarten had moved into the community and not been screened as infants, some had ANSD or may have been missed by the UNHS program. The authors hypothesized that many had late-onset hearing loss and recommended screening programs for older aged children in addition to UNHS. In the current EBSR, included studies performed behavioral audiometric threshold testing at 6–18 months of age, suggesting that, if a second screen is conducted at an older age, it should be considered when a child is still in early childhood.

Conclusions

The results from the current EBSR indicate that newborn hearing screening by ABR or OAEs is often at least moderately effective at identifying permanent hearing loss in early childhood when behavioral audiometric threshold is used as a reference standard. The generalization of these findings to current UNHS programs in the United States is uncertain, as the majority of the included studies were more than 10 years old and many used procedures not currently used in contemporary hearing screening

programs. Additionally, most studies were not conducted in the United States, and it is possible that NICU populations were defined differently.

Previously done EBSRs (IQWiG,2007; USPSTF, 2001; 2008) support that UNHS has effectively lowered the age of identification of hearing loss and that infants who were part of UNHS have higher receptive language outcomes (IQWiG, 2007; USPSTF, 2008). However, no EBSR has evaluated the methodology used to help guide programs in performing best practice. This EBSR was undertaken to provide guidance in developing recommendations for UNHS protocols. The limited results of this EBSR highlight areas that are important for further discussion and future research. First, it is important that, if behavioral measures are to be used as an outcome, future studies provide specific details about the methods and environmental settings used to measure behavioral audiometric thresholds, as well as the personnel performing the hearing evaluations. Second, studies that include babies cared for in the WIN, though costly and cumbersome, should be completed in order to verify screening efficacy in that population and to enable outcomes to be generalized. The results of this EBSR suggest that continued research is needed for UNHS in the United States. Careful attention is needed to assess different screening protocols, stimuli, and response criteria.

Acknowledgments: This EBSR was supported by the American Speech-Language-Hearing Association's National Center for Evidence-Based Practice (N-CEP). We thank the following individuals for their contributions to this manuscript: Rob Mullen, N-CEP director, who served as one of the initial reviewers; N-CEP staff Beverly Wang, who conducted the systematic search, and Rebecca Venediktov, who sifted and appraised the articles; Melissa Wheaton, graduate assistant at Syracuse University, who provided technical assistance; and Anne Oyler, associate director for audiology professional practices, ASHA. The authors have declared no financial or other conflicts in relation to this article.

References

Apostolopoulos, N. K., Psarommatis, I. M., Tsakanikos, M. D., Dellagrammatikas, H. D., & Douniadakis, D. E. (1999). Otoacoustic emission-based hearing screening of a Greek NICU population. *International Journal of Pediatric Otorhinolaryngology, 47*(1), 41–48.

Davis, A., Bamford, J., Wilson, I., Ramkalawan, T., Forshaw, M., & Wright, S. (1997). A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment. *Health Technology Assessment, 1*(10), i–iv, 1–176.

Desai, S., Kollros, P. R., Graziani, L. J., Streletz, L. J., Goodman, M., Stanley, C., et al. (1997). Sensitivity and specificity of the neonatal brain-stem auditory evoked potential for hearing and language deficits in survivors of extracorporeal membrane oxygenation. *Journal of Pediatrics, 131*(2), 233–239.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders.* Baltimore, MD: Brookes.

Durieux-Smith, A., Picton, T. W., Bernard, P., MacMurray, B., & Goodman, J. T. (1991). Prognostic validity of brainstem electric response audiometry in infants of a neonatal intensive care unit. *Audiology, 30*(5), 249–265.

Gaffney, M., Green, D. R., & Gaffney, C. (2010). Newborn hearing screening and follow-up: Are children receiving recommended services? *Public Health Reports, 125,* 199–207. Retrieved from www.cdc.gov/ncbddd/hearingloss/documents/NHS_Follow_Up_2010.pdf.

German Institute for Quality and Efficiency in Health Care (IQWiG). (2007). *Neonatal screening for early detection of hearing impairment. Final report* (S05-01 Version 1.0.). Retrieved from

https://www.iqwig.de/download/S05-01_Final_report_Neonatal_screening_for_early_detection_of_hearing_impairment.pdf.

Gill, A. W., Gosling, D., Kelly, C., Walker, P., & Wooderson, S. (1998). Predischarge screening of very low birthweight infants by click evoked otoacoustic emissions. *Journal of Paediatrics and Child Health, 34*(5), 456–459.

Jacobson, J., & Jacobson, C. (2004). Evaluation of hearing loss in infants and young children. *Pediatric Annals, 33*(12), 811–822.

Joint Committee on Infant Hearing. (2007). Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics, 120*(4), 898–921. Retrieved from http://pediatrics.aappublications.org/cgi/reprint/120/4/898.

Kennedy, C., McCann, D., Campbell, M. J., Kimm, L., & Thornton, R. (2005). Universal newborn screening for permanent childhood hearing impairment: An 8-year follow-up of a controlled trial. *The Lancet, 366*(9486), 660–662.

Kennedy, C. R. (1999). Controlled trial of universal neonatal screening for early identification of permanent childhood hearing impairment: Coverage, positive predictive value, effect on mothers and incremental yield [Wessex Universal Neonatal Screening Trial Group]. *Acta Paediatrica Supplement, 88*(432), 73–75.

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics, 106*(3), E43.

Nelson, H. D., Bougatsos, C., & Nygren, P. (2008). Universal newborn hearing screening: systematic review to update the 2001 U.S. Preventive Services Task Force recommendation. *Pediatrics, 122*(1), e266–276.

Norton, S. J., Gorga, M. P., Widen, J. E., Folsom, R. C., Sininger, Y., Cone-Wesson, B., et al. (2000). Identification of neonatal hearing impairment: evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance. *Ear and Hearing, 21*(5), 508–528.

Prieve, B. A. (2000). Identification of neonatal hearing impairment: A cornerstone for newborn hearing screening. *Ear and Hearing, 21*(5), 345.

Prieve, B. A., & Stevens, F. (2000). The New York State universal newborn hearing screening demonstration project: Introduction and overview. *Ear and Hearing, 21*(2), 85–91.

Roth, D. A., Hildesheimer, M., Bardenstein, S., Goidel, D., Reichman, B., Maayan-Metzger, A., et al. (2008). Preauricular skin tags and ear pits are associated with permanent hearing impairment in newborns. *Pediatrics, 122*(4), e884–890.

Savio, G., Perez-Abalo, M. C., Gaya, J., Hernandez, O., & Mijares, E. (2006). Test accuracy and prognostic validity of multiple auditory steady state responses for targeted hearing screening. *International Journal of Audiology, 45*(2), 109–120.

Shimizu, H., Walters, R. J., Proctor, L. R., Kennedy, D. W., Allen, M. C., & Markowitz, R. K. (1990). Identification of hearing impairment in the neonatal intensive care unit population: Outcome of a five-year project at the Johns Hopkins Hospital. *Seminars in Hearing, 11*(2), 150–160.

Sininger, Y. S., Grimes, A., & Christensen, E. (2010). Auditory development in early amplified children: Factors influencing auditory-based communication outcomes in children with hearing loss. *Ear and Hearing, 31*(2), 166–185.

Smyth, V., Scott, J., & Tudehope, D. (1990). The utility of the auditory brainstem response as a screening procedure. *International Journal of Pediatric Otorhinolaryngology, 19*(1), 45–55.

Stevens, J. C., Webb, H. D., Hutchinson, J., Connell, J., Smith, M. F., & Buffin, J. T. (1990). Click evoked otoacoustic emissions in neonatal screening. *Ear and Hearing, 11*(2), 128–133.

Swigonski, N., Shallop, J., Bull, M. J., & Lemons, J. A. (1987). Hearing screening of high risk newborns. *Ear and Hearing, 8*(1), 26–30.

U.S. Preventive Services Task Force. (2001). Newborn hearing screening: recommendations and rationale. *American Family Physician, 64*(12), 1995–1999.

U.S. Preventive Services Task Force. (2008). *Universal screening for hearing loss in newborns* [Recommendation statement]. Retrieved from www.uspreventiveservicestaskforce.org/uspstf08/newbornhear/newbhearrs.htm.

Watkin, P. M. & Baldwin, M. (2011). Identifying deafness in early childhood: Requirements after the newborn hearing screen. *Archives of Disease in Childhood, 96*, 62–66.

Watkin, P. M., Baldwin, M., & McEnery, G. (1991). Neonatal at risk screening and the identification of deafness. *Archives of Disease in Childhood, 66*(10 Spec No), 1130–1135.

Widen, J. E., Folsom, R. C., Cone-Wesson, B., Carty, L., Dunnell, J. J., Koebsell, K., et al. (2000). Identification of neonatal hearing impairment: Hearing status at 8 to 12 months corrected age using a visual reinforcement audiometry protocol. *Ear and Hearing, 21*(5), 471–487.

Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K., & Mehl, A. L. (1998). Language of early- and later-identified children with hearing loss. *Pediatrics, 102*(5), 1161–1171.

Table 1. Quality Indicators for Included Studies.

Indicator	Quality Indicator
Study design	∗ *Prospective.* • Retrospective.
Reference standard	∗ *Appropriate/reasonable reference standard used for comparison.* • Reference standard not appropriate or reasonable for comparison.
Selection/recruitment	* *Random or consecutive selection.* • Convenience sample or hand-picked sample or not stated.
Blinding	∗ *Assessors blinded when interpreting results of test and reference.* • Assessors not blinded when interpreting results of test and reference or not stated.
Participants	∗ *Participants adequately described and similar to population in which tests would be used with full spectrum of severity.* • Participants not adequately described or participants not similar to population in which test would be used with full spectrum of severity.
Avoidance of verification bias	∗ *Reference standard given to all participants.* • Reference standard not given to all participants but decision to perform reference standard independent of test results. • Reference standard not given to participants and decision to perform reference standard not independent of test results or not stated.
Likelihood ratios	∗ *Likelihood ratios reported or calculable.* • Likelihood ratios neither reported nor calculable.
Follow-up *(prospective studies only)*	∗ *Results reported on all subjects entered into study.* • Reasonable loss to follow up, ≤ 20% of results not reported. • Greater than 20% of results not reported.

**Italicized quality marker indicates highest level of quality.*

Table 2. Interpretation of LR.

LR	Interpretation
> 10	Large and often conclusive increase in the likelihood of disorder
5–10	Moderate increase in the likelihood of disorder
2–5	Small increase in the likelihood of disorder
1–2	Minimal increase in the likelihood of disorder
1	No change in the likelihood of disorder
0.5–1.0	Minimal decrease in the likelihood of disorder
0.2–0.5	Small decrease in the likelihood of disorder
0.1–0.2	Moderate decrease in the likelihood of disorder
< 0.1	Large and often conclusive decrease in the likelihood of disorder

Source: http:// *http://omerad.msu.edu/ebm/Diagnosis/Diagnosis6.html*

Table 3. Description of Participants.

Citation	Number Screened	Number Ears Screened	Age Range at Initial Screening	Nursery Placement At Birth and Participant Characteristics	Number Followed-Up	Number of Ears Followed-Up	Age Range at Follow-Up	Type, Degree and # of Participants Or Ears With PCHL
Apostolopoulos et al. (1999)	223	438	72 hours -28 days Mean = 19.2 days	WIN: 0 NICU: 223	107	213	> 2 1/2 years	3 SNHL, unspecified degree
Ari-Even Roth et al. (2008)	637	NR	Infancy (screened before discharge from birth admission)	NR Pits and tags	151	NR	7–36 months	15 total: 5 SNHL: 1 bilateral mild-moderate, 3 bilateral moderate, 1 unilateral severe 8 conductive: 1 bilateral mild, 2 bilateral moderate, 1 bilateral moderate-severe, 4 unilateral moderate-severe. 2 mixed: Both unilateral severe-profound
Desai et al. (1997)	80	NR	< 1 month	WIN: 0 NICU: 80 ECMO	80	NR	10–12 months 18–24 months 30–48 months	12 SNHL, unspecified degree: 3 unilateral 9 bilateral
Durieux-Smith et al. (1991)	600	NR	infancy (presumed <1 year)	WIN: 0 NICU: 600	333	NR	3 years for study	13 SNHL: 6 unilateral 7 bilateral
Gill et al. (1998)	144	NR	24–37 weeks gestational age Median 29 weeks	WIN: 0 NICU: 144 All VLBW	87	NR	>10 months	10 SNHL
Norton et al. (2000)	4911	NR	< 1 day to > 10 days WIN, not sure for NICU	80, 353 with JCIH risk factor NICU: 4478	3134 followed, 2995 successful	5554 successful ears, 301 partial ears, 135 no data	8–12 months	86 ears (56 infants) with permanent hearing loss (30 infants with bilateral hearing loss): 26 mild 21 moderate 18 severe 21 profound

Savio et al. (2006)	508	NR	3 months corrected age	WIN: 0 NICU: 508	125	NR	4–5 months corrected age; again at 3–4 years	17 total: 15 SNHL (1 unilateral) 2 bilateral mixed
Shimizu et al (1990)	458	NR	31–53 weeks PCA, Mean age 39.1 weeks	WIN: 0 NICU: 348	338	NR	18 months	7 total: 4 moderate 2 severe 1 unspecified (subject 229)
Smyth et al. (1990)	149	NR	31–49 weeks gestational age mean = 39.93 weeks	WIN: 0 NICU: 149	133	NR	About 7 months	1, unspecified degree
Stevens et al. (1990)	723	NR	Mean post conceptional age when tested 37.5 weeks for inpatients and 47 weeks for outpatients Range = 32–49 weeks	NR: reports "mostly NICU"	331	NR	≥ 8 months Corrected age	4 or 5 unspecified hearing loss
Swigonski et al. (1987)	137 or 138 (Inconsistently reported)	172	Mean PCA = 36.9 weeks Range = 32–48 weeks	WIN: 0 NICU: 137 or 138 (inconsistently reported)	82 or 83 (inconsistently reported)	NR	6 and 9 months	4 severe

Note: ECMO – extracorporeal membrane oxygenation; NICU – neonatal intensive care unit; NR – not reported; PCA – post-conceptual age; PCHL – permanent childhood hearing loss; SNHL – sensorineural hearing loss; WIN – well-infant nursery; VLBW – very low birthweight.

Table 4. Quality Appraisal for all Studies.

Diagnostic Studies

Citation	Study Design	Reasonable Reference Standard Used?	Selection or Recruitment	Blinding	Subjects Similar to Population in Which Test Is Performed?	Avoidance of Verification Bias	Likelihood Ratios Reported or Calculable?	Follow-Up
Apostolopoulos et al. (1999)	Prospective	Yes	Random or consecutive selection	Not stated	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Ari-Even Roth et al. (2008)	Retrospective	Yes	Convenience sample/Hand-picked sample	Not stated	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Desai et al. (1997)	Prospective	Yes	Convenience sample/Hand-picked sample	Yes	No	Reference standard given to all subjects	Yes	Results reported on all participants
Durieux-Smith et al. (1991)	Prospective	Yes	Not stated	Yes	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Gill et al. (1998)	Prospective	Yes	Not stated	Not stated	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Norton et al. (2000)	Prospective	Yes	Random or consecutive selection	Not stated	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Savio et al. (2006)	Prospective	Yes	Convenience sample/Hand-picked sample	Not stated	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Shimizu et al. (1990)	Prospective	Yes	Random or consecutive selection	Not stated	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Smyth et al. (1990)	Prospective	Yes	Convenience sample/Hand-picked sample	Not stated	No	Reference standard not given to all subjects but decision to perform reference standard independent of test results	Yes	≤ 20% of results not reported
Stevens et al. (1990)	Prospective	Yes	Convenience sample/Hand-picked sample	Yes	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Swigonski et al. (1987)	Prospective	Yes	Convenience sample/Hand-picked sample	Yes	No	Reference standard given to all subjects	Yes	> 20% of results not reported
Watkin et al. (1991)	Prospective	Yes	Convenience sample/Hand-picked sample	Not stated	No	Reference standard given to all subjects	Yes	≤ 20% of results not reported

Note: Shaded areas indicate highest level of quality in each category

Table 5. OAE Results.

Citation	Screening Tool(s)	Stimulus Parameters and Response Criteria	Reference Standard	Definition of HL	Sensitivity	Specificity	Likelihood Ratios	Quality Marker score
Apostolopoulos et al. (1999)	TEOAEs (ABR)	ILO88 FullScreen. 80 µs clicks 75–85 dB peSPL, ≥ 80% stability, at least 100 samples in average. Pass was SNR of 3 dB or better in 3 bands: 1–2, 2–3 and 3–4 kHz	Behavioral testing using BOA, VRA and CPA when child was > 2.5 yrs of age.	NR	90.9%	91.1%	LR+: 10.21 LR-: 0.01	5/8
Ari-Even Roth et al. (2008)	TEOAEs	NR	Behavioral testing, elevated behavioral threshold	> 25 dB HL: 500–4000 Hz	100%	71%	LR+: 3.33 LR-: 0	3/8
Gill et al. (1998)	TEOAEs	ILO Quickscreen	VROA	NR	85%	87%	LR+: 6.35 LR-: 0.18	4/8
Norton et al. (2000)	DPOAEs (ABR)	L1: L2 = 65/50 dB SPL f2 = 1, 1,5, 2, 3, 4, kHz; f2/f1 = 1.22 Stop criteria: SNR ≥ 3dB higher than 2SDs above mean noise	Ear-specific VRA at 1, 2, 4 kHz and SAT. Did not test lower than 20 dB HL. VRA tested at 8–12 months corrected age.	MRLs: ≥ 30 dB HL	PTA 2&4: 88%	PTA 2&4: 83%	LR+: 5.2 LR-: 0.14	5/8
	DPOAEs	L1:L2 = 75/75 dB SPL			PTA 2&4: 78%	PTA 2&4: 82%	LR+: 4.3 LR-: 0.27	
	TEOAEs (ABR)	80 dB pSPL Custom click Stop criteria SNR in 4/5, ½-octave bands: 3 dB SNR at 1 &1.5 kHz; 6 dB SNR at 2, 3, & 4 kHz.			PTA 2&4: 83%	PTA 2&4: 90%	LR+: 8.3 LR-: 0.19	
Stevens et al. (1990)	TEOAEs (ABR)	Custom system. Click 100µs rarefaction at 32.5/s. Nonlinear trace obtained by 31/41 or 41/51 dB nHL. Scored present or absent by two independent scorers.	All babies entered were recalled at 8 months for "distraction testing and tympanometry at 8 months corrected age". "full head turn to a range of stimuli covering the audiometric frequencies"	(a) > 30 dB nHL for better ear or (b) > 40 dB nHL	OAE (a): 55% OAE (b): 67%	OAE (a): 82% OAE (b): 82%	LR+: 3.1 LR-: 0.54 LR+: 3.72 LR-: 0.4	5/8

ABR – Auditory brainstem response; BOA – behavioral response audiometry; CPA – conditioned play audiometry; dB nHL – dB normal hearing level; dB HL – dB hearing level; DPOAEs – Distortion-product otoacoustic emissions; LR+ – positive likelihood ratio; LR- – negative likelihood ratio; MRL – minimal response level; NR – not reported; PTA – pure tone average; SAT – Speech awareness threshold; SNR – signal-to-noise ratio; TEOAEs – transient-evoked otoacoustic emissions; VRA – visual reinforcement audiometry; VROA – visual reinforcement orientation audiometry

Table 6. ABR and ASSR Results.

ABR Studies

Citation	Screening Tool(s)	Stimulus and Response Parameters	Behavioral Reference Standard	Definition of HL	Sensitivity	Specificity	Likelihood Ratios	Quality Marker Score
Desai et al. (1997)	CABR	100 µs clicks at 11.5/s through TDH 39 earphones. Fail when Wave V at levels of 45 or 85 dB nHL or I-V interval > mean +2.5 SDs of norms by visual interpretation.	Click ABR and behavioral audiometry done at several ages for confirmation, starting at 10–12 mo of age and ending at 30–48 mo of age.	25 dB HL from 250 to4000 Hz	42%	76%	LR+: 1.75 LR-:0.76	6/8
Durieux-Smith et al. (1991)	CABR	100 µs rarefaction clicks at 61/s through TDH 49 earphones at 30 dB nHL (re: 10 adults w/ threshold @ 40 dB pSPL). Visual interpretation.	Pure tone testing at 3 years - 0.5, 1, 2, 4 kHz, immittance.	25 dBHL from 500 to 4000 Hz	All Hearing Loss 43.3% All SNHL and Mixed 61.5% Bilateral SNHL and Mixed Requiring Amplification 86%	93.6% 99.3% 100%	LR+: 6.77 LR-: 0.61 LR+: 87.86 LR-: 0.39 LR+: NC LR-: 0.14	5/8
Norton et al. (2000)	SABR (OAE)	Click at 30 dB nHL through OAE probe. Stop criteria: Fsp=2.4 (20% of data) and Fsp=3.1 (80% of data).	Ear-specific VRA at 1, 2, 4 kHz and SAT. Did not test lower than 20 db HL. VRA tested at 8-12 months corrected age.	MRLs≥30 dB HL	PTA2&4kHz: 82%	PTA2&4 kHz: 90%	LR+: 8.2 LR-: 0.2	5/8
Savio et al. (2006)	CABR (ASSR)	100 µs click, 40 dB nHL(ref = 75 dB pSPL) through TDH 49earphone at 17/s. Visual identification Wave V.	1st f/u: SF testing, otoscopy, immittance, reflexes. Also MSSR at 0 .5, 1, 2, & 4kHz and CABR 2nd f/u: complete behavioral audiometry and speech/lang screen (ELM)	MRLs≥25 dB HL	94% 71.3%		LR+: 3.28 LR-: 0.84	4/8
Shimizu et al. (1990)	CABR	Clicks at 22/sec through TDH39 earphone at 30 & 70 dB nHLwith 37 out of the 88 failures also tested at 40. Pass was at 30 dB nHL (0 nHL = 25 dBpeSPL). Two observers judged responses.	VRA at 18 months, and ABR if needed. CPA or VRA at 3–4 years.	Considered normal if thresholds were ≤25 dB by air conduction from 500 to 4000 Hz OR SRT at 25 dB or lower with a normal tympanogram.	100%	77.1%	LR+: 2.359 LR-:0	5/8

Smyth et al. (1990)	CABR	Clicks presented at 20/s,TDH49 earphones hand-held. Testing in shielded test suite. Wave V by visual inspection of thresholds, morphology and latency of Wave V. Passing level is less than or equal to 40 dB nHL.	VROA using the conditioned orienting response (COR). Warble tones, pure tones and speech stimuli presented in sound field with 10 dB step sizes. For some subjects, used monaural earphones but not all (not specified how many).	Considered normal if thresholds were ≤ 40 dB HL.	64%	58%	LR+: 2.359 LR-: 0.63	3/8
Stevens et al. (1990)	CABR (OAE)	Clicks, 100 µs, throughTDH39 earphone at 32.5/s. Pass was 43 dBnHL in one ear and 53 dBnHL in other ear by visual inspection, agreement of two observers.	All babies recalled at 8 months corrected age for distraction testing and tympanometry. Required "full head turn to a range of stimuli covering the audiometric frequencies."	(a) ≤30 dBnHL for better ear for normal (b) ≤40 dBnHL in better ear for normal	ABR (a): 45% ABR (b): 67%	ABR(a): 91% ABR(b): 91%	LR+: 5 LR-: 0.6 LR+: 7.44 LR-: 0.36	5/8
Swigonski et al. (1987)	CABR	Click125 us duration, alternating polarity at 20/s, through TDH-49, hand-held. Visual inspection by two observers: 40 HL = pass, 60 = conditional, 80 = fail. Our calculations based on conditional passes were fails.	COR in SF at 500 and 4000 Hz, warble-tone or NBN. Testing done at 6 and 9 months	6 months: normal was 25–30 dB HL; 9 mos, normal was 15–20 dB HL.	100%	70%	LR+: 3.33 LR-: 0	5/8
Watkin et al. (1991)	CABR	Alternating clicks at 50?s, TDH-39 earphone, hand-held Two repeatable waveforms by visual inspection at 40 dB nHL in one ear was a pass.	Infant distraction test, admittance, otoscopy by audiologist at 7 mo. If an infant did not attend the follow up, the health visitor's distraction test was obtained, (did not indicate # of cases).	Bilateral, at least moderate in better ear.	100%	87%	LR+: 7.69 LR-: 0	4/8
ASSR Study								
Savio et al. (2006)	ASSR (ABR)	40 dB nHL (ref = 62 dB SPL RMS) through TDH 49 earphones. Used multiple frequency stimuli at 500 & 2000 Hz (MSSR) with depths of 95% at 95 and 101 Hz, respectively. MSSR detected automatically based on Hotelling T2 test ($p<0.05$ at each frequency).	1st f/u: complete audio, otocopy, immitance, reflexes, MSSR at .5, 1, 2, & 4kHz, cABR, and SF behavioral testing 2nd f/u: complete behavioral audio w/ speech/lang screen (ELM)	MRLs>25 dB HL	100%	71.3%	LR+: 3.48 LR-: 0	4/8

ABR - Auditory brainstem response; ASSR - auditory steady state response; BOA – behavioral response audiometry; CABR – conventional ABR; COR - conditioned orienting response; CPA – conditioned play audiometry; ELM – Early Language Milestones; dB nHL – dB normal hearing level; dB HL – dB hearing level; LR+ - positive likelihood ratio; LR- - negative likelihood ratio; MRL – minimal response level; MSSR: multiple auditory steady state responses; NBN - narrow band noise; NC – not calculable; NR – not reported; OAEs – otoacoustic emissions; PTA – pure tone average; SABR – screening ABR; SAT – Speech awareness threshold; SF- sound field; SNR – signal-to-noise ratio; VRA – visual reinforcement audiometry; VROA – visual reinforcement orientation audiometry.

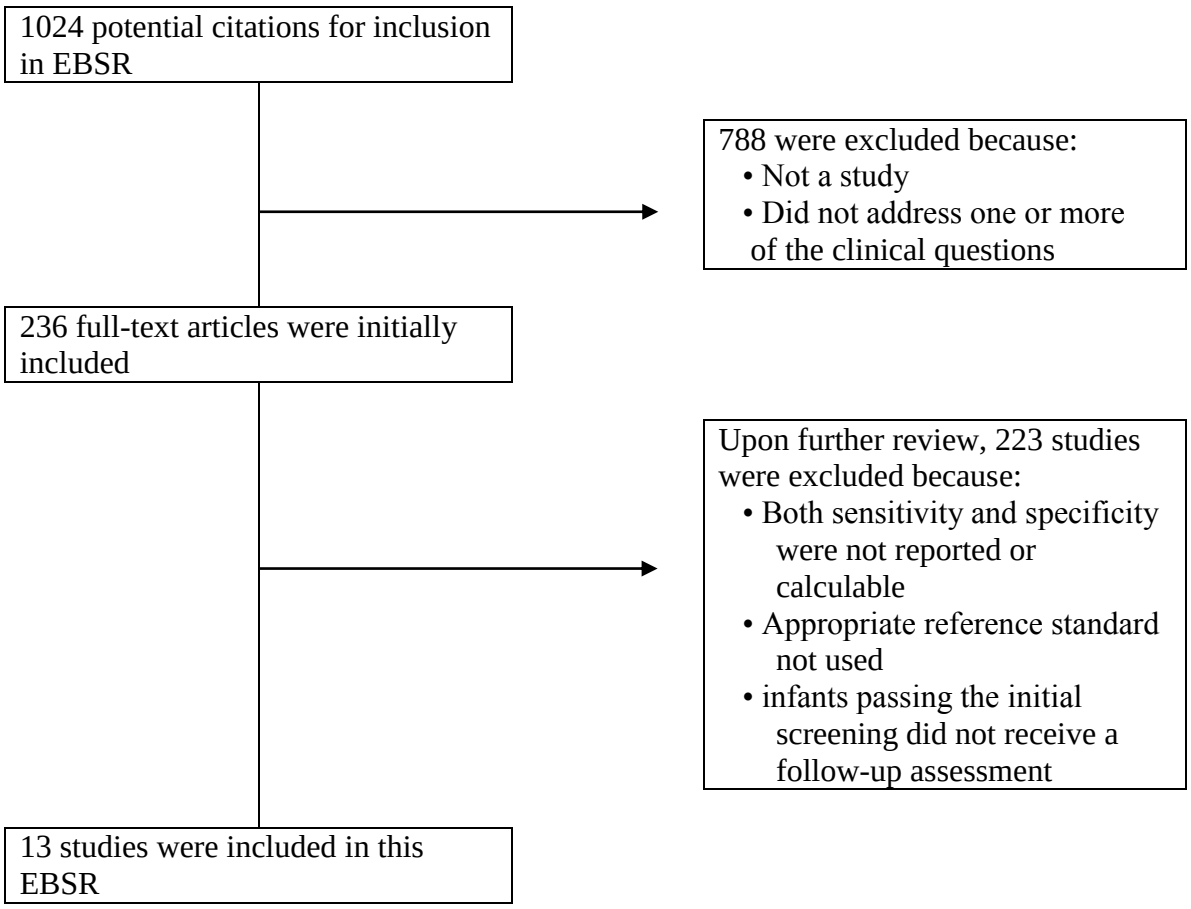
Table 7. Effectiveness of Different OAE Criteria.

TEOAE				DPOAE 65/50				DPOAE 75/75			
Criteria		**PT 2+4 kHz**	**PT 1, 2+4 kHz**	**Criteria**		**PT 2+4 kHz**	**PT 1, 2+4 kHz**	**Criteria**		**PT 2+4 kHz**	**PT 1, 2+4 kHz**
dB SNR				**dB SNR**				**dB SNR**			
3	Sens	85	98	**3**	sens	87	92	**3**	sens	72	88
	Spec	88	42		spec	85	45		spec	92	60
	LR+	7.08	1.69		LR+	5.80	1.67		LR+	9.00	2.20
	LR-	0.17	0.05		LR-	0.15	0.18		LR-	0.30	0.20
6	Sens	88	98	**6**	sens	91	100	**6**	sens	80	100
	Spec	85	25		spec	45	8		spec	70	12
	LR+	5.87	1.31		LR+	1.65	1.09		LR+	2.67	1.14
	LR-	0.14	0.08		LR-	0.20	0.00		LR-	0.29	0.00
9	Sens	90	98	**9**	sens	98	100	**9**	sens	88	100
	Spec	58	12		spec	28	5		spec	58	8
	LR+	2.14	1.11		LR+	1.36	1.05		LR+	2.10	1.09
	LR-	0.17	0.17		LR-	0.07	0.00		LR-	0.21	0.00

Values presented in this table were extracted from Norton et al., (2000).

Note: DPOAEs - distortion product otoacoustic emissions; LR+ - positive likelihood ratio; LR - negative likelihood ratio; OAEs - otoacoustic emissions; PT - pure-tone average; sens - sensitivity; spec - specificity; SNR - signal-to-noise ratio; TEOAEs - transient-evoked otoacoustic emissions.

*Figure 1***.** Process for Identification of Included Studies.



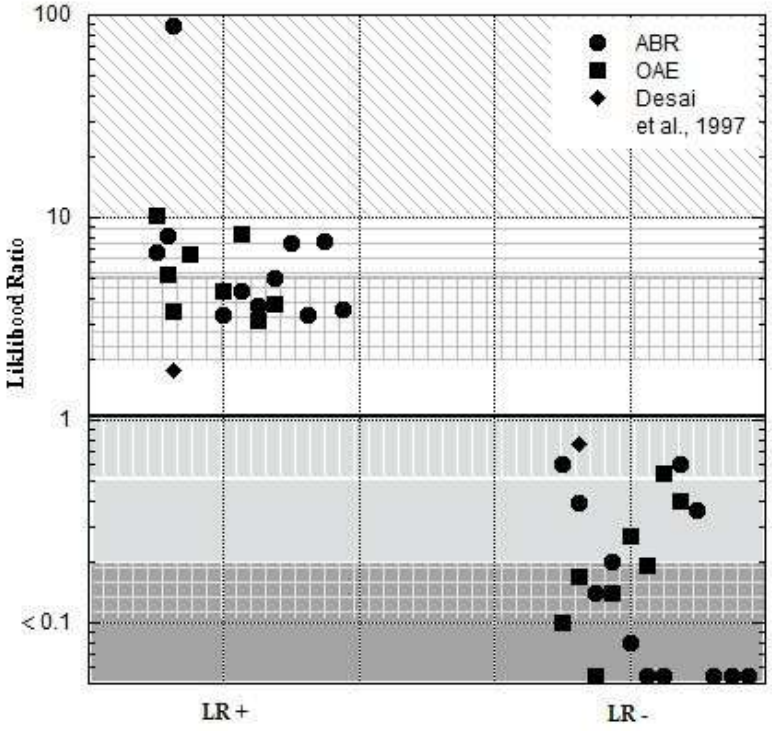


Figure 2. LR+ and LR- for all studies included in Tables 7 and 8. LRs from OAE studies are represented by circles and LRs from ABR studies are represented as squares. LRs from Desai et al., 1997, a study on babies that underwent ECMO, are represented by diamonds. The shading on the figure corresponds to LR interpretation as given in Table 1.